A simple framework to justify linear response theory

February 17, 2010

Martin Hairer and Andrew J. Majda

Courant Institute, New York University
Email: mhairer@cims.nyu.edu, Email: jonjon@cims.nyu.edu

Abstract

The use of linear response theory for forced dissipative stochastic dynamical systems through the fluctuation dissipation theorem is an attractive way to study climate change systematically among other applications. Here, a mathematically rigorous justification of linear response theory for forced dissipative stochastic dynamical systems is developed. The main results are formulated in an abstract setting and apply to suitable systems, in finite and infinite dimensions, that are of interest in climate change science and other applications.

1 Introduction

One of the cornerstones of modern statistical physics is the fluctuation dissipation theorem (FDT) which roughly states that for systems of identical particles in statistical equilibrium, the average response to small external perturbations can be calculated through the knowledge of suitable correlation functions of the unperturbed statistical system. For some of the many practical applications of the FDT, see for example [KTH85, Bal97]. The low frequency response to changes in external forcing for various components of the climate system is a central problem of contemporary climate change science. Leith [Lei75] suggested that if the climate system satisfied a suitable FDT, then the climate response to small external forcing should be calculated by estimating suitable statistics in the present climate. The climate system is a forced dissipative chaotic dynamical system which is physically quite far from the classical physicists' setting for FDT. Leith's suggestion has created a lot of recent activity in generating new theoretical formulations and approximate algorithms for FDT that were applied to climate response with considerable skill ([GD99, GBD02, GB07, GBM08, AM07, AM08, AM09, MW10]).

The goal here is to provide a rigorous justification of linear response theory at infinite times for a rather large class of general stochastic systems, including a class of problems with potential applications to climate change science. The improved regularity due to the presence of noise simplifies key aspects of the problem considerably and circumvents the difficulties found to occur even in relatively simple deterministic systems (see [Rue97, Bal08] for an overview and [BY93, KL99, GL06, GL08, BS08] for a number of advances in this direction). As a consequence, we are able to focus on systems that have some other features of realistic systems, namely a lack of ellipticity (we will only require hypoellipticity), non-compactness of state space, a lack of global Lipschitz continuity of the coefficients, and even infinite-dimensionality of the state

space. The results established below apply to a very broad class of nonlinear functionals of state space which trivially include many quantities of interest, such as the mean and variance of subsets of variables.

The mathematical formulation of FDT as linear response theory for forced dissipative stochastic dynamical systems [Ris89, MAG05, MFK08, Pal01, MW10] is an appropriate setting for these applications since many current improvements in the comprehensive computer models for climate change involve stochastic components [PW10], while lower dimensional reduced models typically also involve stochastic noise terms [MFK08, MAG10, MGY10]. While there is a recent formal systematic discussion of FDT for time-dependent stochastic systems of interest in climate change science [MW10], it is not obvious whether the results of this article generalise in a straightforward way. The remainder of this article is structured as follows. After introducing our setup and notations, a general abstract formulation of the main Theorem is presented in Section 2 with a general discussion on its applicability provided in Section 3. Section 4 illustrates the scope of the theorem for finite-dimensional forced dissipative stochastic systems with a prototype structure which is relevant for some applications in climate change science.

Acknowledgements

2   Main result

Throughout the entire article, we will consider a family of Markov evolutions $\{\mathcal{P}_t^a : t \geq 0, \, a \in \mathbf{R}\}$ on a separable (either finite or infinite-dimensional) Banach space $\mathcal{B}$. In other words, the $\mathcal{P}_t^a$ are Markov operators over $\mathcal{B}$ such that $\mathcal{P}_0^a \varphi = \varphi$ and $\mathcal{P}_t^a \mathcal{P}_s^a \varphi = \mathcal{P}_{s+t}^a \varphi$ for every bounded measurable function $\varphi \colon \mathcal{B} \to \mathbf{R}$. In this abstract setting, we take the pragmatic view that a Markov operator $\mathcal{P}$ is determined by a transition kernel P as

$$(\mathcal{P}\varphi)(x) = \int_{\mathcal{B}} \varphi(y) \, P(x, dy) \,,$$

where the map $A \mapsto P(x, A)$ yields a Borel probability measure for every $x \in \mathcal{B}$ and that the map $x \mapsto P(x, A)$ is measurable for every Borel set A. We will denote by $\mathcal{P}^*$ the adjoint of $\mathcal{P}$ acting on the space of finite signed Borel measures on $\mathcal{B}$ by

$$(\mathcal{P}^*\mu)(A) = \int_{\mathcal{B}} P(y, A) \, \mu(dy) \,.$$

We will also fix a particular value $a_0 \in \mathbf{R}$ and we are interested in the long-time behaviour of the system described by $\mathcal{P}_t^a$ for a close to a_0. Throughout this article, we assume that for every a, there exists a probability measure μ_a on $\mathcal{B}$ which is invariant for $\mathcal{P}_t^a$, so that $(\mathcal{P}_t^a)^* \mu_a = \mu_a$ for every $t \geq 0$. The aim of this note is to show that, under very weak natural assumptions on the dynamic, the map $a \mapsto \langle \varphi, \mu_a \rangle$ is differentiable at a_0 for every sufficiently regular test function $\varphi \colon \mathcal{B} \to \mathbf{R}$. (Note that the fact that $a \in \mathbf{R}$ is not a restriction since if we take $a \in \mathbf{R}^n$ instead, it suffices to consider its components one by one, keeping the others fixed.)

One of the main features of our proof will be that it relies on rather 'soft' arguments and has conditions that are easily verifiable in practice. For example, it does *not* require the transition probabilities to have a density with respect to some reference measure (typically Lebesgue in finite dimensions or Gaussian in infinite dimensions).

Denote by $\mathcal{C}_0^\infty(\mathcal{B})$ the set of all functions $\varphi \colon \mathcal{B} \to \mathbf{R}$ such that there exists $N > 0$, a linear map $T \colon \mathcal{B} \to \mathbf{R}^N$, and a smooth, compactly supported map $\hat{\varphi} \colon \mathbf{R}^N \to \mathbf{R}$ such that $\varphi = \hat{\varphi} \circ T$. Given two continuous functions $V, W \colon \mathcal{B} \to [1, \infty)$, we set $\mathcal{C}_{V,W}^1$ to be the closure of $\mathcal{C}_0^\infty(\mathcal{B})$ under the norm

$$\|\varphi\|_{1;V,W} = \sup_{x \in \mathcal{B}} \Big(\frac{|\varphi(x)|}{V(x)} + \frac{\|D\varphi(x)\|}{W(x)} \Big) . \tag{2.1}$$

If we quotient this space by the space of all constant functions, then this norm is equivalent to the Lipschitz norm corresponding to the distance function

$$d_{V,W}(x, y) = \varrho_W(x, y) \wedge (V(x) + V(y)) . \tag{2.2}$$

Here, for any positive continuous function W bounded away from 0, we define the metric

$$\varrho_W(x, y) = \inf_{\gamma(0)=x;\gamma(1)=y} \int_0^1 W(\gamma(s)) \|\dot{\gamma}(s)\| \, ds ,$$

where the infimum runs over all differentiable curves $\gamma \colon [0, 1] \to \mathcal{B}$ with the prescribed boundary conditions. (Note that if $W \equiv 1$, then one simply has $\varrho_W(x, y) = \|x - y\|$.) We then make the following assumption:

Assumption 1 *There exists a time $t > 0$ and a constant $\varrho < 1$ such that $\|\mathcal{P}_t^{a_0}\varphi - \langle \varphi, \mu_{a_0} \rangle\|_{1;V,W} \leq \varrho \|\varphi - \langle \varphi, \mu_{a_0} \rangle\|_{1;V,W}$ for every $\varphi \in \mathcal{C}_{V,W}^1$.*

Remark 2.1 This spectral gap condition is quite different from the L^2 or L^p spectral gap conditions often encountered in the literature. It is however possible to verify that such a condition is satisfied for a very large class of finite-dimensional and infinite-dimensional (hypo-)elliptic diffusions. See [HM08] for the verification of Assumption 1 in the case of the 2D stochastic Navier-Stokes equations and Section 4 below for a class of finite-dimensional SDEs.

In particular, Assumption 1 implies that the invariant measure for $\mathcal{P}_t^{a_0}$ is unique, that the spectrum of $\mathcal{P}_t^{a_0}$ as an operator on $\mathcal{C}_{V,W}^1$ has an isolated eigenvalue of multiplicity 1 at 1, and that the remainder of its spectrum is contained in the ball of radius ϱ around the origin. Therefore, for every centred function $\varphi \in \mathcal{C}_{V,W}^1$, there exists a unique centred function $\psi \in \mathcal{C}_{V,W}^1$ such that $\psi - \mathcal{P}_t^{a_0}\psi = \varphi$. We will henceforth use the notation $\psi = (1 - \mathcal{P}_t^{a_0})^{-1}\varphi$.

Our second assumption concerns the regularity of $\mathcal{P}_t^a$ with respect to the parameter a. In order to state this assumption, we introduce the space $\mathcal{C}_U$, which is the weighted space of continuous functions obtained by completing $\mathcal{C}_0^\infty(\mathcal{B})$ under the norm

$$\|\varphi\|_U = \sup_{x \in \mathcal{B}} \frac{|\varphi(x)|}{U(x)} .$$

We then assume that

Assumption 2 *There exists a continuous function $U \geq V$ from $\mathcal{B}$ to $[1, \infty)$ such that, for some fixed $t > 0$ and every $\varphi \in \mathcal{C}^1_{V,W}$, the map $a \mapsto \mathcal{P}^a_t \varphi$ is differentiable in a neighbourhood of a_0 when viewed as a map from $\mathbf{R}$ to $\mathcal{C}_U$. Denoting this derivative by $P^a_{\cdot t} \varphi$, we furthermore assume that*

$$\|P^a_{\cdot t} \varphi\|_U \leq C \|\varphi\|_{1;V,W} , \tag{2.3}$$

for some constant $C > 0$.

Remark 2.2 Note how we are allowed to lose one derivative as well as some growth in the norm of φ when computing $P^a_{\cdot t} \varphi$. This is what makes the verification of (2.3) very easy in practice, provided that U, V, and W are chosen in a judicious way, see Section 4 below. When considering such problems on spaces of probability measures instead of spaces of test functions, it has been know for some time that in the context of dynamical systems, spaces of $\mathcal{C}^1$ functions were suitable to take advantage of the expanding character of a map [Rue89]. Several highly non-trivial extensions of this fact were considered much more recently in [GL06, BL07, BS08]. See also [AGY06] for an instance of the use of weights in a related dynamical systems context.

Finally, we assume that we have an *a priori* bound on the integrability of the invariant measures:

Assumption 3 *There exists $\varepsilon > 0$ such that $\sup_{|a-a_0|<\varepsilon} \int_{\mathcal{B}} U(x)\, \mu_a(dx) < \infty$.*

Our main result then states that:

Theorem 2.3 *Let $\{\mathcal{P}^a_t\}_{a \in \mathbf{R}}$ be a family of Markov semigroups over a separable Banach space $\mathcal{B}$ such that there exist $\mathcal{C}^1$ functions $U, V, W : \mathcal{B} \to \mathbf{R}_+$ such that Assumptions 1, 2 and 3 are satisfied for some $a = a_0$ and some fixed $t > 0$. Then the map $a \mapsto \langle \varphi, \mu_a \rangle$ is differentiable at a_0 for every $\varphi \in \mathcal{C}^1_{V,W}$ and the identity*

$$\frac{d}{da}\langle \varphi, \mu_a \rangle \Big|_{a=a_0} = \langle P^{a_0}_{\cdot t}(1 - \mathcal{P}^{a_0}_t)^{-1}(\varphi - \langle \varphi, \mu_{a_0} \rangle), \mu_{a_0} \rangle , \tag{2.4}$$

holds. In particular, the right hand side of (2.4) is well-defined.

Remark 2.4 As pointed out by one of the referees, the spectral gap condition is not essential for Theorem 2.3 to hold. It could be replaced by the requirement that the map $a \mapsto \mu_a$ is continuous in the topology of weak convergence and that the test function φ is such that there exists a function $\psi \in \mathcal{C}^1_{V,W}$ such that $(1 - \mathcal{P}^{a_0}_t)\psi = \varphi - \langle \varphi, \mu_{a_0} \rangle$.

It is not clear how easy the latter condition would prove to verify in practice for systems that do not possess a spectral gap as it would require a detailed understanding of the properties of the resolvent which is bypassed by Assumption 1.

Remark 2.5 The formula (2.4) is identical to that familiar to dynamicists, see for example [Bal08, Eq. 3]. Since the left hand side in (2.4) is independent of t, we can take the limit $t \to 0$ in the right hand side. If the test function φ is centred with respect to μ_{a_0} and sufficiently 'nice' so that $t(1 - \mathcal{P}^{a_0}_t)^{-1}\varphi \to \mathcal{L}^{-1}_{a_0}\varphi$, we then obtain the linear response formula

$$\frac{d}{da}\langle \varphi, \mu_a \rangle \Big|_{a=a_0} = \langle \mathbf{L}_{\cdot a_0} \mathcal{L}^{-1}_{a_0}\varphi, \mu_{a_0} \rangle , \tag{2.5}$$

where L_a is the formal derivative of the generator $\mathcal{L}_a$ with respect to the parameter a. (This can also be obtained by formally differentiating the identity $\mathcal{L}_a^* \mu_a = 0$ with respect to the parameter a.)

If the state space is $\mathbf{R}^n$, μ_{a_0} has a density $\exp(-H(x))$ with respect to Lebesgue measure and $\mathsf{L}_{a_0} = F(x)\nabla_x$ is a first-order differential operator, then we can define the conjugate current J by

$$J(x) = F(x)\nabla_x H(x) - \operatorname{div}_x F(x) .$$

With this notation, (2.5) can be interpreted as the Green-Kubo relations

$$\frac{d}{da}\langle \varphi, \mu_a \rangle \Big|_{a=a_0} = \langle J\mathcal{L}_{a_0}^{-1}\varphi, \mu_{a_0} \rangle = \int_0^\infty \mathbf{E}(J(x_0)\varphi(x_t))\, dt ,$$

where expectations are taken over the stationary Markov process x_t with invariant measure μ_{a_0}.

Remark 2.6 Note that even though we assume uniqueness of the invariant measure for $\mathcal{P}_t^{a_0}$, this need not be the case for $a \neq a_0$. Even though uniqueness doesn't need to hold for $a \neq a_0$, our result implies differentiability of the map $a \mapsto \langle \varphi, \mu_a \rangle$ at $a = a_0$ for any choice of $a \mapsto \mu_a$. This is analogous to the result shown for partially hyperbolic systems in [Dol04].

Remark 2.7 The classical perturbation results found for example in [Kat76] do not seem to be applicable directly to our situation since we do *not* assume that the map $a \mapsto \mathcal{P}_t^a \varphi$ is differentiable in $\mathcal{C}_{V,W}^1$.

Proof. Let us first show that the map $a \mapsto \mu_a$ is Lipschitz continuous at $a = a_0$ if we endow the space of probability measures on $\mathcal{B}$ which integrate $d_{V,W}(0, x)$ with the norm dual to that of $\mathcal{C}_{V,W}^1$. For every $\varphi \in \mathcal{C}_{V,W}^1$, we have the identity

$$\langle \varphi, \mu_a - \mu_{a_0} \rangle = \langle \mathcal{P}_t^{a_0}\varphi, \mu_a - \mu_{a_0} \rangle + \langle (\mathcal{P}_t^a - \mathcal{P}_t^{a_0})\varphi, \mu_a \rangle .$$

Since none of these expressions changes if we add a constant to φ, we can assume in the sequel that φ is centred with respect to μ_{a_0} so that, using the invertibility of $(1 - \mathcal{P}_t^{a_0})$ and setting $\psi = (1 - \mathcal{P}_t^{a_0})\varphi$, we obtain the identity

$$\langle \psi, \mu_a - \mu_{a_0} \rangle = \langle (\mathcal{P}_t^a - \mathcal{P}_t^{a_0})(1 - \mathcal{P}_t^{a_0})^{-1}\psi, \mu_a \rangle . \tag{2.6}$$

It now follows readily from Assumptions 2 and 3 that

$$\sup_{\|\varphi\|_{1;V,W}=1} \langle \varphi, \mu_a - \mu_{a_0} \rangle = \sup_{\|\varphi\|_{1;V,W}=1} \langle (\mathcal{P}_t^a - \mathcal{P}_t^{a_0})(1 - \mathcal{P}_t^{a_0})^{-1}\varphi, \mu_a \rangle$$

$$\leq C|a - a_0| \sup_{\|\psi\|_U=1} \langle \psi, \mu_a \rangle \leq \tilde{C}|a - a_0| ,$$

for some positive constants C and $\tilde{C}$, as required.

Since the norm dual to that of $\mathcal{C}_{V,W}^1$ (restricted to differences between probability measures) is nothing but the Wasserstein-1 distance corresponding to $d_{V,W}$ [Vil03], it follows that $\mu_a \to \mu_{a_0}$ weakly for $a \to a_0$. Combining this with Assumption 3, we conclude in particular that

$$\langle \Psi, \mu_a \rangle \to \langle \Psi, \mu_{a_0} \rangle , \qquad \forall \Psi \in \mathcal{C}_U . \tag{2.7}$$

(This follows by a standard 'diagonal argument' from the fact that Ψ can be approximated by continuous bounded functions by the definition of $\mathcal{C}_U$.)

We now have all the preliminaries in place to show that $\langle \varphi, \mu_a \rangle$ is differentiable and to compute its derivative. It follows from Assumption 2 that, for every (centred) $\varphi \in \mathcal{C}^1_{V,W}$, there exists a function $a \mapsto \mathcal{R}_a \in \mathcal{C}_U$ with $\lim_{a \to a_0} \|\mathcal{R}_a\|_U = 0$ and such that we have the identity

$$\frac{1}{a - a_0}(\mathcal{P}^a_t - \mathcal{P}^{a_0}_t)(1 - \mathcal{P}^{a_0}_t)^{-1}\varphi = \dot{\mathrm{P}}^{a_0}_t(1 - \mathcal{P}^{a_0}_t)^{-1}\varphi + \mathcal{R}_a \ .$$

As a consequence of (2.6), we have

$$\begin{aligned}
\langle \varphi, \mu_a - \mu_{a_0} \rangle &= \langle (\mathcal{P}^a_t - \mathcal{P}^{a_0}_t)(1 - \mathcal{P}^{a_0}_t)^{-1}\varphi, \mu_{a_0} \rangle \\
&\quad + \langle (\mathcal{P}^a_t - \mathcal{P}^{a_0}_t)(1 - \mathcal{P}^{a_0}_t)^{-1}\varphi, \mu_a - \mu_{a_0} \rangle \ ,
\end{aligned}$$

so that

$$\begin{aligned}
\left| \frac{\langle \varphi, \mu_a - \mu_{a_0} \rangle}{a - a_0} - \langle \dot{\mathrm{P}}^{a_0}_t(1 - \mathcal{P}^{a_0}_t)^{-1}\varphi, \mu_{a_0} \rangle \right| &\leq 2|\langle \mathcal{R}_a, \mu_{a_0} \rangle| + |\langle \mathcal{R}_a, \mu_a \rangle| \\
&\quad + |\langle \dot{\mathrm{P}}^{a_0}_t(1 - \mathcal{P}^{a_0}_t)^{-1}\varphi, \mu_a - \mu_{a_0} \rangle| \ .
\end{aligned}$$

The first two terms on the right hand side converge to zero thanks to the continuity of $\mathcal{R}_a$ at a_0 combined with Assumption 3, while the last term converges to zero thanks to (2.7), thus concluding the proof. $\qquad\square$

3 Domain of applicability

In this section, we discuss under which circumstances one can expect a spectral gap result in a norm of the type (2.1). Harris' theorem [HM09b, MT93] provides a general methodology to verify whether a system satisfies a spectral gap in $\mathcal{C}_V$ for some V. If it so happens that such a system also has the property that one can find W such that, for some fixed $t_0 > 0$, $\mathcal{P}^{a_0}_{t_0}$ is a bounded operator from $\mathcal{C}_V$ into $\mathcal{C}^1_{V,W}$, then it follows that $\mathcal{P}^{a_0}_t$ has a spectral gap in $\mathcal{C}^1_{V,W}$. However, such an approach presents two problems:

1. While Harris' theorem is suitable for finite-dimensional diffusions, there are many examples of stochastic PDEs satisfying a spectral gap in a space of the type $\mathcal{C}^1_{V,W}$ for which Harris' theorem does not apply.

2. In finite dimensions, as soon as a system satisfies Hörmander's bracket condition, $\mathcal{P}^{a_0}_t$ maps continuous functions into smooth functions. In particular, there always exists some W such that it maps $\mathcal{C}_V$ into $\mathcal{C}^1_{V,W}$. The problem is that there does not appear to be a general result available that controls the growth of such a function W. This is problematic, since the function U appearing in Assumption 2 is generally comparable to W, so that Assumption 3 is hard to verify without control on W.

Our aim is therefore to present conditions under which a spectral gap in a space of the type $\mathcal{C}^1_{V,W}$ can be verified directly, without requiring a spectral gap result in a space of the type $\mathcal{C}_V$. In [HMS09], a general framework was laid out that gives conditions under which a "spectral gap" is obtained in a Wasserstein-type distance. Unfortunately, the Wasserstein distance in question is with respect to a metric comparable to the square root of the original metric on the space. Therefore, in our setting, this actually yields

a spectral gap in a space of $\frac{1}{2}$-Hölder continuous functions, which is useless for our purpose. On the other hand, a spectral gap in a space of the type (2.1) was obtained for the stochastic Navier-Stokes equations in [HM08].

In this section, we obtain a slightly different set of conditions under which a spectral gap in a norm of the type (2.1) can be shown for a general Markov operator $\mathcal{P}$ over a Banach space $\mathcal{B}$. Our conditions are inspired by those in [HM08, HMS09], aiming in particular at being applicable to equations with conservative quadratic nonlinearities, as arising in applications. The main ingredient is again a gradient bound of the following type, which was shown to hold for a large class of stochastic PDEs in [HM06, HM09a]:

Assumption 4 *There exist continuous functions* $U_1, U_2 \colon \mathcal{B} \to \mathbf{R}_+$ *such that, for every* $\varepsilon > 0$ *there is a constant* $C_\varepsilon > 0$ *such that the bound*

$$\|D\mathcal{P}\varphi(x)\|^2 \le \varepsilon U_1^2(x)(\mathcal{P}\|D\varphi\|^2)(x) + C_\varepsilon U_2^2(x)(\mathcal{P}\varphi^2)(x)\,, \qquad (3.1)$$

holds for every $x \in \mathcal{B}$ *and every* $\varphi \in \mathcal{C}_b^1(\mathcal{B})$.

Remark 3.1 In the case of the stochastic 2D Navier-Stokes equations, this bound can be shown to hold with $U_1(x) = U_2(x) = \exp(\eta\|x\|^2)$ for arbitrarily small $\eta > 0$, with x the vorticity of the vector field and $\|x\|$ its L^2-norm.

The aim of this section is to show that if a Markov operator $\mathcal{P}$ satisfies Assumption 4 and has sufficiently good contraction properties, then it is possible to deduce a spectral gap in the norm (2.1) (for a suitable choice of V and W), provided that the transition probabilities satisfy a kind of topological irreducibility condition. This statement can be formulated precisely in the following way:

Theorem 3.2 *Let* $\mathcal{P}$ *be a Markov operator over a separable Banach space* $\mathcal{B}$ *mapping* $\mathcal{C}_0^\infty(\mathcal{B})$ *into* $\mathcal{C}_{V,W}^1$ *and satisfying Assumption 4. Suppose furthermore that there exist continuous functions* $V, W \colon \mathcal{B} \to \mathbf{R}_+$ *satisfying*

$$U_1^2 \mathcal{P}W^2 + U_2^2 \mathcal{P}V^2 \le CW^2\,, \qquad \mathcal{P}V \le \frac{1}{2}V + K\,, \qquad (3.2)$$

for some constants C *and* K.

Finally, assume that there exists a point $x_\star \in \mathcal{B}$ *such that, for every* $\varepsilon > 0$ *and every* $C > 0$ *there exists* $\alpha > 0$ *such that*

$$\inf_{x\,\colon V(x)\le C} \mathcal{P}(x, B_\varepsilon(x_\star)) \ge \alpha\,. \qquad (3.3)$$

Then, $\mathcal{P}$ *has exactly one invariant probability measure* $\mu_\star$. *Furthermore, there exist constants* C *and* $\gamma > 0$ *such that the bound*

$$\|\mathcal{P}^n\varphi - \langle\varphi, \mu_\star\rangle\|_{1;V,W} \le Ce^{-\gamma n}\|\varphi - \langle\varphi, \mu_\star\rangle\|_{1;V,W}\,, \qquad (3.4)$$

holds for every $\varphi \in \mathcal{C}_{V,W}^1$. *Finally, there exist constants* $\delta > 0$, $\beta > 0$ *and* $\varrho < 1$ *such that the bound*

$$\|\mathcal{P}\varphi - \langle\varphi, \mu_\star\rangle\|_{1;1+\beta V, \delta^{-1}W} \le \varrho\|\varphi - \langle\varphi, \mu_\star\rangle\|_{1;1+\beta V, \delta^{-1}W}\,, \qquad (3.5)$$

holds for every $\varphi \in \mathcal{C}_{V,W}^1$.

Remark 3.3 It will follow from the proof that it is sufficient that the second inequality in (3.2) holds with $\frac{1}{2}$ replaced by any other constant $\kappa < 1$. Furthermore, we do not need (3.3) to hold for every $C > 0$, but only for some $C > K/(1 - \kappa)$.

Proof. Since the norms $\| \cdot \|_{1;1+\beta V,\delta^{-1}W}$ and $\| \cdot \|_{1;V,W}$ are equivalent for every choice of $\beta, \delta > 0$, (3.4) follows immediately from (3.5), so that it is sufficient to check (3.5).

The proof uses very similar ideas to the proofs of [HMS09, Theorem 4.7] and [HM08, Theorem 3.4]. In particular, we use the following trick. For $\delta > 0$ and $\beta > 0$, we introduce the distance

$$\hat{d}_{\beta,\delta}(x, y) = \delta^{-1} \varrho_W(x, y) \wedge (2 + \beta V(x) + \beta V(y)) .$$

This distance is of course equivalent to the distance $d_{V,W}$ introduced in (2.2), but it turns out that allowing the freedom of choosing both δ and β sufficiently small will considerably simplify the proofs. With this definition, a $\mathcal{C}^1$ function φ is Lipschitz continuous with Lipschitz constant 1 with respect to $\hat{d}_{\beta,\delta}$ if and only if

$$\|D\varphi(x)\| \leq \delta^{-1}W(x) , \qquad |\varphi(x)| \leq 1 + \beta V(x) . \tag{3.6}$$

(For the second inequality, one might have to add a suitable constant to φ.) Denote by $\mathrm{Lip}_{\delta,\beta}$ the set of all such functions. As in [HMS09, HM08], we show that it is possible to choose δ and β in such a way that the bound

$$\hat{d}_{\beta,\delta}(\mathcal{P}(x, \cdot), \mathcal{P}(y, \cdot)) \leq \alpha d_{\beta,\delta}(x, y) , \tag{3.7}$$

holds for some $\alpha < 1$ uniformly over all pairs $x, y \in \mathcal{B}$. We again show (3.7) separately in three different cases and we use separately the three ingredients of the theorem in each of these cases.

The case $\varrho_W(x, y) \leq \delta(2 + \beta V(x) + \beta V(y))$. In this case, we make use of the gradient bound (3.1), together with the 'super-Lyapunov' structure (3.2) to deduce that if φ satisfies (3.6), then for every $\varepsilon > 0$ there exists C_ε such that the bound

$$\|D\mathcal{P}\varphi(x)\| \leq \varepsilon\delta^{-1}W(x) + C_\varepsilon W(x) ,$$

holds uniformly for all such φ and for all $\beta \leq 1$, say. It follows that by first choosing $\varepsilon = \frac{1}{4}$ and then choosing δ small enough so that $C_\varepsilon \leq 1/(4\delta)$, one has

$$\|D\mathcal{P}\varphi(x)\| \leq \frac{1}{2\delta}W(x) ,$$

which immediately implies that

$$\begin{aligned}
\hat{d}_{\beta,\delta}(\mathcal{P}(x, \cdot), \mathcal{P}(y, \cdot)) &\leq \sup_{\varphi \in \mathrm{Lip}_{\delta,\beta}} |\mathcal{P}\varphi(x) - \mathcal{P}\varphi(y)| \\
&\leq \sup_{\varphi \in \mathrm{Lip}_{\delta,\beta}} \inf_\gamma \int_0^1 \|D\mathcal{P}\varphi(\gamma(s))\| \, |\dot{\gamma}(s)| \, ds \\
&\leq \frac{1}{2\delta} \sup_\varphi \inf_\gamma \int_0^1 W(\gamma(s)) \, |\dot{\gamma}(s)| \, ds \\
&\leq \frac{1}{2\delta} \varrho_W(x, y) \leq \frac{1}{2}\hat{d}_{\beta,\delta}(x, y) ,
\end{aligned}$$

as requested.

The case $\varrho_W(x,y) > \delta(2+\beta V(x)+\beta V(y))$ and $V(x)+V(y) \geq 4(K+2)$.
In this case, we only make use of the fact that V is a Lyapunov function. We have
indeed

$$\hat{d}_{\beta,\delta}(\mathcal{P}(x,\cdot),\mathcal{P}(y,\cdot)) \leq \sup_{\varphi \in \mathrm{Lip}_{\delta,\beta}} |\mathcal{P}\varphi(x) - \mathcal{P}\varphi(y)| \leq \sup_{\varphi \in \mathrm{Lip}_{\delta,\beta}} (|\mathcal{P}\varphi(x)| + |\mathcal{P}\varphi(y)|)$$

$$\leq 2 + \beta\mathcal{P}V(x) + \beta\mathcal{P}V(y) \leq 2 + \frac{\beta}{2}(V(x)+V(y)) + \beta K$$

$$\leq (2-2\beta) + \frac{3\beta}{4}(V(x)+V(y)) \leq (1-(\beta \wedge \tfrac{1}{4}))\hat{d}_{\beta,\delta}(x,y) \,,$$

which again yields a contraction, but with a strength that depends this time on the
parameter β. Finally, we have

The case $\varrho_W(x,y) > \delta(2+\beta V(x)+\beta V(y))$ and $V(x)+V(y) < 4(K+2)$.
In this case, we make use of our final assumption, namely (3.3). At this stage we
assume that $\delta > 0$ is fixed, sufficiently small so that our first step goes through. We
can then find some sufficiently small $\varepsilon > 0$ so that $\hat{d}_{\beta,\delta}(x_\star, y) \leq \frac{1}{2}$ for all $y \in B_\varepsilon(x_\star)$,
uniformly over $\beta \leq 1$. In this case, we can decompose the Markov operator $\mathcal{P}$ into a
combination $\mathcal{P} = \alpha\mathcal{P}_1 + (1-\alpha)\mathcal{P}_2$ of Markov operators such that $\mathcal{P}_1(x, B_\varepsilon(x_\star)) = 1$
for every x such that $V(x) \leq 4(K+2)$. We conclude that

$$\hat{d}_{\beta,\delta}(\mathcal{P}(x,\cdot),\mathcal{P}(y,\cdot)) \leq \alpha\hat{d}_{\beta,\delta}(\mathcal{P}_1(x,\cdot),\mathcal{P}_1(y,\cdot)) + (1-\alpha)\hat{d}_{\beta,\delta}(\mathcal{P}_2(x,\cdot),\mathcal{P}_2(y,\cdot))$$

$$\leq \alpha + (1-\alpha)(2 + \beta\mathcal{P}_2 V(x) + \beta\mathcal{P}_2 V(x))$$

$$\leq \alpha + 2(1-\alpha) + \beta\mathcal{P}V(x) + \beta\mathcal{P}V(x)$$

$$\leq 2 - \alpha + \frac{\beta}{2}(V(x)+V(y)+2K) \leq 2 - \alpha + \beta(3K+4) \,.$$

We can now choose β sufficiently small so that this constant is strictly smaller than 2.
Since on the other hand one has $\hat{d}_{\beta,\delta}(x,y) \geq 2$, the claim now follows. $\square$

4 Application to stochastic differential equations

In this section, we aim to apply the results of the previous sections to a concrete class
of stochastic differential equations. In order to keep our conditions clean, we restrict
ourselves to SDEs with additive noise and polynomial nonlinearities. We thus assume
that for some integers $N, M \geq 1$, we have

$$dx(t) = \sum_{k=0}^{N} \mathcal{N}_k(x,\ldots,x)\, dt + \sum_{k=1}^{M} \sigma_k\, dw_k(t) \,, \qquad x(0) \in \mathbf{R}^d \,. \tag{4.1}$$

Here, the functions $\mathcal{N}_k$ are symmetric multilinear maps of order k and the σ_k are ele-
ments of $\mathbf{R}^d$. We also define the function $\mathcal{N} \colon \mathbf{R}^d \to \mathbf{R}^d$ by

$$\mathcal{N}(x) = \sum_{k=0}^{N} \mathcal{N}_k(x,\ldots,x) \,,$$

as well as the $d \times M$ matrix Σ given by $\Sigma = (\sigma_1,\ldots,\sigma_M)$.

When $N = 2$, equations of this type typically arise as effective dynamic for large-
scale structures in climate models [MAG05]. Fixing N and M, we make the following
structural assumptions on the $\mathcal{N}_k$'s and the σ_k's:

Assumption 5 *There exist constants $c, C > 0$ such that the bound*

$$\langle x, \mathcal{N}(x) \rangle \leq C - c\|x\|^N , \quad \forall x \in \mathbf{R}^d \tag{4.2}$$

holds.

Remark 4.1 The bound (4.2) can hold even if the degree N of the nonlinearity is even. In this case, it implies that $\langle x, \mathcal{N}_N(x, \dots, x) \rangle = 0$. If N is odd, then (4.2) would actually also hold with N replaced by $N + 1$, but this is not important for our purpose.

Assumption 6 *Define $A_k \subset \mathbf{R}^d$ recursively by $A_0 = \text{span}\{\sigma_1, \dots, \sigma_M\}$ and*

$$A_{k+1} = \text{span}(A_k \cup \{\mathcal{N}_N(y_1, \dots, y_N) : y_j \in A_k\}) .$$

Then, one has $A_d = \mathbf{R}^d$.

Assumption 7 *The control system associated to (4.1) is approximately controllable. In other words, for every initial condition x_0, every $T > 0$, every terminal condition x_T, and every $\varepsilon > 0$, there exists a smooth control $u \in \mathcal{C}^\infty([0, T], \mathbf{R}^M)$ such that the solution to*

$$\dot{x}(t) = \mathcal{N}(x(t)) + \sigma\, u(t) , \qquad x(0) = x_0 ,$$

satisfies $\|x(T) - x_T\| \leq \varepsilon$.

Remark 4.2 If N is odd, then Assumption 7 is redundant since it then follows from Assumption 6, see [JK85]. However, in the more interesting case where N is even, the situation is not so easy. A number of criteria are given in [Sus87, Kaw90, MBB03], but they are all sufficient without being necessary. In the particular case of truncations of the two and three-dimensional Navier-Stokes equations where furthermore Σ is assumed to be diagonal in the Fourier basis, it was also shown that Assumption 7 follows from Assumption 6 [EM01, Rom04, AS05].

For given $M, N > 0$, let $\mathscr{J}$ denote the set of $(\mathcal{N}, \Sigma)$ such that Assumptions 5, 6, and 7 hold. Note that $(\mathcal{N}, \Sigma)$ can be viewed as a vector in $\mathbf{R}^n$ with $n = d(M + \sum_{k=0}^N \binom{d+k}{k}))$, so that $\mathscr{J}$ is nothing but some possibly rather complicated open subset of $\mathbf{R}^n$. With these assumptions in hand, we have the following result, which follows immediately from [DPZ96, MT93]:

Theorem 4.3 *For every $(\mathcal{N}, \Sigma) \in \mathscr{J}$, the system (4.1) possesses a unique invariant probability measure $\mu_{(\mathcal{N}, \Sigma)}$.*

Proof. The existence of an invariant measure follows from Assumption 5 by Krylov-Bogoliubov. The uniqueness is then a consequence of the strong Feller property (which follows from Assumption 6) and the topological irreducibility (which follows from Assumption 7) of the associated Markov semigroup. $\square$

The aim of this section is to prove the following result:

Theorem 4.4 *Fix $M, N \geq 1$ and let $\varphi \colon \mathbf{R}^d \to \mathbf{R}$ be a $\mathcal{C}^1$ function such that*

$$\sup_{x \in \mathbf{R}^d} (|\varphi(x)| + \|D\varphi(x)\|)e^{-\delta\|x\|^2} < \infty ,$$

for every $\delta > 0$. Then, the map $(\mathcal{N}, \Sigma) \mapsto \int_{\mathbf{R}^d} \varphi(x)\, \mu_{(\mathcal{N}, \Sigma)}(dx)$ is $\mathcal{C}^1$ on $\mathscr{J}$ and its derivative is given by the formula (2.4).

Before we turn to the proof of Theorem 4.4, we start by providing the *a priori* bounds on the solutions to (4.1) required for our analysis. For this, we first introduce the Jacobian process of (4.1), which is the $d \times d$ matrix-valued two-parameter process $J_{s,t}$ given by the solution to the following random ODE:

$$\frac{dJ_{s,t}}{dt} = D\mathcal{N}(x)J_{s,t} \ , \qquad J_{s,s} = \text{Id} \ . \tag{4.3}$$

We also introduce the second variation process $J_{s,t}^{(2)}$ which is a bilinear map-valued process given by

$$\frac{dJ_{s,t}^{(2)}(\xi,\zeta)}{dt} = D\mathcal{N}(x)J_{s,t}^{(2)}(\xi,\zeta) + D^2\mathcal{N}(x)(J_{s,t}\xi, J_{s,t}\zeta) \ , \qquad J_{s,s}^{(2)} = 0 \ ,$$

for every ξ and ζ in $\mathbf{R}^d$. We then have the following:

Proposition 4.5 *Let $(\mathcal{N}, \Sigma)$ be such that Assumption 5 holds. Then, there exist strictly positive constants η, ν and C and, for every $\varepsilon > 0$ and $p > 0$ there exists a constant $C_{p,\varepsilon}$ such that the bounds*

$$\mathbf{E} \exp\Big(\eta\|x(t)\|^2 + \eta \int_0^t \|x(s)\|^N \, ds \Big) \leq C \exp(\eta e^{-\nu t}\|x(0)\|^2) \ , \tag{4.4}$$

$$\mathbf{E}(\|J_{s,t}\|^p + \|J_{s,t}^{(2)}\|^p) \leq C_{p,\varepsilon} \exp(\varepsilon\|x(0)\|^2) \ , \tag{4.5}$$

hold for every initial condition $x(0) \in \mathbf{R}^d$, and for every $s, t \in [0,1]$.

Proof. Applying Itô's formula to the function $x \mapsto \|x\|^2$, it follows from our assumption that

$$\frac{1}{2}d\|x(t)\|^2 \leq -c\|x\|^N \, dt + C \, dt + \sum_{k=1}^M \langle x(t), \sigma_k \rangle \, dw_k(t) \ ,$$

for a possibly different constant C. Note that this inequality should really be interpreted in integral form. The bound (4.4) then follows from the version of the exponential martingale inequality given in [HM08, Lemma 5.1].

In order to obtain the bound (4.5), note that it follows from (4.3) that there exists a constant C such that

$$\|J_{s,t}\|^p \leq \exp\Big(pC \int_s^t (1 + \|x(r)\|^{N-1}) \, dr \Big) \ .$$

The requested bound now follows immediately from (4.4), noting that for every $\delta > 0$ there exists a constant $\tilde{C}_{p,\delta}$ such that the inequality

$$pC(1 + \|x\|^{N-1}) \leq \delta\|x\|^N + \tilde{C}_{p,\delta} \ ,$$

holds for every $x \in \mathbf{R}^d$. □

The main ingredient for all probabilistic proofs of regularising properties for SDEs is the *Malliavin matrix* of the process. This is the $\mathbf{R}^{d \times d}$-valued matrix $\mathcal{M}_t$ defined by

$$\mathcal{M}_t = \int_0^t J_{s,t}\Sigma\Sigma^*J_{s,t}^* \, ds \ ,$$

where $J_{s,t}$ is the Jacobian of the flow as above and Σ is the noise matrix defined at the start of this section. Note that $\mathcal{M}_t$ is a random object which furthermore depends on the initial condition x_0 of (4.1). We have the following *a priori* bound:

Proposition 4.6 *Let $(\mathcal{N}, \Sigma)$ be such that Assumptions 5 and 6 hold. Then, for every $t > 0$, every $\delta > 0$, and every $p \geq 1$ there exists a constant $C > 0$ such that the bound*

$$\mathbf{E}\|\mathcal{M}_t^{-\frac{1}{2}}\|^p \leq C \exp(\delta\|x_0\|^2) \,,$$

holds for every $x_0 \in \mathbf{R}^d$.

Proof. This is an immediate consequence of [HM09a, Thm 6.7]. The assumptions of that theorem are satisfied by Proposition 4.5 if we choose $\psi_0(x) = \exp(\eta\|x\|^2)$ for a sufficiently small constant $\eta > 0$. □

As a consequence of Proposition 4.6, we obtain the following gradient bound for the Markov semigroup $\mathcal{P}_t$ generated by solutions to (4.1):

Proposition 4.7 *Let $(\mathcal{N}, \Sigma)$ be such that Assumptions 5 and 6 hold. Then, for every $t > 0$ and every $\eta > 0$ there exists constant C such that the bound*

$$\|D\mathcal{P}_t\varphi(x)\|^2 \leq C \exp(\eta\|x\|^2)(\mathcal{P}_t\varphi^2)(x) \,,$$

holds for every $x \in \mathbf{R}^d$ and every $\varphi \in \mathcal{C}_b^1(\mathbf{R}^d)$.

Proof. The proof is almost identical to that of [HM09a, Thm 5.4], which is in turn largely inspired by that of [Nor86, Thm 3.2]. We therefore only give a sketch of the proof and refer to [Nor86, HM09a, Nua06] for details. Let $\xi \in \mathbf{R}^d$ be an arbitrary deterministic vector with $\|\xi\| = 1$ and fix a terminal time $t > 0$. We then define the $\mathbf{R}^M$-valued stochastic process h_ξ by

$$h_\xi(s) = \Sigma^* J_{s,t}^* \mathcal{M}_t^{-1} J_{0,t}\xi \,.$$

Note that t is a *fixed* terminal time, so that h_ξ is a stochastic process which is not adapted to the filtration generated by the Wiener processes driving (4.1). With this notation at hand, it follows from the integration by parts formula on Wiener space [Nua06] that one has the identity

$$\langle D\mathcal{P}_t\varphi(x), \xi\rangle = \mathbf{E}\Big(\varphi(x_t)\int_0^t h_\xi(s)\,dW(s)\Big) \,,$$

where the stochastic integral should be interpreted as a Skorokhod integral [Nua06]. It then follows from the generalised Itô isometry [Nua06] that

$$\langle D\mathcal{P}_t\varphi(x), \xi\rangle^2 \leq (\mathcal{P}_t\varphi^2)(x)\,\mathbf{E}\Big(\int_0^t \|h_\xi(s)\|^2\,ds + \int_0^t \|\mathscr{D}_r h_\xi(s)\|^2\,ds\,dr\Big) \,,$$

where $\mathscr{D}_r F$ denotes the Malliavin derivative at time r of the random variable F. It follows from the chain rule that one has the identity

$$\mathscr{D}_r h_\xi(s) = \Sigma^* \mathscr{D}_r J_{s,t}^* \mathcal{M}_t^{-1} J_{0,t}\xi + \Sigma^* J_{s,t}^* \mathcal{M}_t^{-1} \mathscr{D}_r J_{0,t}\xi$$

$$- \int_0^t \Sigma^* J_{s,t}^* \mathcal{M}_t^{-1}(\mathscr{D}_r J_{u,t}\Sigma\Sigma^* J_{u,t}^* + J_{u,t}\Sigma\Sigma^* \mathscr{D}_r J_{u,t}^*)\mathcal{M}_t^{-1} J_{0,t}\xi\,du \,.$$

Note now that the Malliavin derivative of the Jacobian $J_{s,t}$ is given by

$$\mathscr{D}_r J_{s,t}\xi = \begin{cases} 0 & \text{if } r \geq t, \\ J_{s,t}^{(2)}(J_{r,s}\Sigma, \xi) & \text{if } r \leq s, \\ J_{r,t}^{(2)}(\Sigma, J_{s,r}\xi) & \text{otherwise.} \end{cases}$$

Collecting all of these identities, it then follows immediately from the a priori bounds of Proposition 4.5 that the requested bound holds. □

With all of these preliminary results at hand, we can now proceed to the proof of Theorem 4.4:

Proof of Theorem 4.4. As a first step, we show that the assumptions of Theorem 3.2 hold. It follows from Proposition 4.7 that Assumption 4 is satisfied with $U_1(x) = 0$ and $U_2(x) = \exp(\delta\|x\|^2)$ for arbitrary choice of $\delta > 0$. In view of Proposition 4.5, it is therefore natural to choose $V(x) = W(x) = \exp(\eta\|x\|^2)$ for some sufficiently small value of $\eta > 0$. It is possible to first choose η small enough so that the bounds in Proposition 4.5 hold and then δ small enough so that the bounds (3.2) hold.

We now want to show that (3.3) holds as well. It follows from Assumption 7 combined with the support theorem [SV79] that $\mathcal{P}_t(x, B_\varepsilon(y)) > 0$ for every $x, y \in \mathbf{R}^d$, $t > 0$ and $\varepsilon > 0$. Since furthermore the Markov semigroup $\mathcal{P}_t$ is Feller and the level sets of V are compact, (3.3) follows at once. We thus conclude that, for every $t > 0$, the assumptions of Theorem 3.2 hold, so that the semigroup $\mathcal{P}_t$ possesses a spectral gap in the space $\mathcal{C}^1_{V,V}$ for sufficiently small values $\eta > 0$.

We now show that Theorem 2.3 can be applied to our situation, if we choose a to be any of the components of either $\mathcal{N}$ or Σ. First of all, it follows from our previous discussion that Assumption 1 is satisfied for every $t > 0$, provided that we choose $V(x) = 1 + \beta\exp(\eta\|x\|^2)$ and $W(x) = \delta^{-1}\exp(\eta\|x\|^2)$ for sufficiently small values β, δ, and η (possibly depending on the value of t). Furthermore, if we set $U(x) = 2\exp(\eta'\|x\|^2)$ for some sufficiently small value $\eta' > \eta$, then it follows from (4.4) that Assumption 3 is also satisfied. It therefore suffices to check that Assumption 2 is also satisfied.

Denote by $\Phi_t(x, \mathcal{N}, \Sigma)$ the solution to (4.1) with initial condition x and parameters $\mathcal{N}$ and Σ. By considering $\mathcal{N}$ and Σ as additional dynamical variables that follow the evolution equations $\dot{\mathcal{N}} = 0$, $\dot{\Sigma} = 0$, it follows from [Kun90, CE83] that on $\mathbf{R}^d \times \mathscr{J}$, Φ_t is a $\mathcal{C}^\infty$ map of all of its arguments. Furthermore, it follows from the variation of constants formula that its derivative with respect to $\mathcal{N}$ and Σ in arbitrary directions $\delta\mathcal{N}$ and $\delta\Sigma$ is given by

$$\langle D_{\mathcal{N}}\Phi_t, \delta\mathcal{N}\rangle = \int_0^t J_{s,t}\delta\mathcal{N}(x_s)\,ds\,, \quad \langle D_{\Sigma}\Phi_t, \delta\Sigma\rangle = \int_0^t J_{s,t}\delta\Sigma\,dW(s)\,.$$

Combining this with the bounds from Proposition 4.5, it follows at once that Assumption 2 is satisfied with

$$V(x) = 1 + \beta e^{\eta\|x\|^2}\,, \quad W(x) = \delta^{-1}e^{\eta\|x\|^2}\,, \quad U(x) = V(x) + e^{2\eta\|x\|^2}\,,$$

provided that β, δ, and η are sufficiently small. This allows us to apply Theorem 2.3, thus concluding the proof. □

References

[AGY06] A. AVILA, S. GOUËZEL, and J.-C. YOCCOZ. Exponential mixing for the Teichmüller flow. *Publ. Math. Inst. Hautes Études Sci.* **104**, (2006), 143–211.

[AM07] R. V. ABRAMOV and A. J. MAJDA. Blended response algorithms for linear fluctuation-dissipation for complex nonlinear dynamical systems. *Nonlinearity* **20**, no. 12, (2007), 2793–2821.

[AM08] R. V. ABRAMOV and A. J. MAJDA. New approximations and tests of linear fluctuation-response for chaotic nonlinear forced-dissipative dynamical systems. *J. Nonlinear Sci.* **18**, no. 3, (2008), 303–341.

[AM09]   R. V. ABRAMOV and A. J. MAJDA. A new algorithm for low frequency climate response. *J. Atmos. Sci.* **66**, no. 2, (2009), 286–309.

[AS05]   A. A. AGRACHEV and A. V. SARYCHEV. Navier-Stokes equations: controllability by means of low modes forcing. *J. Math. Fluid Mech.* **7**, no. 1, (2005), 108–152.

[Bal97]   R. BALESCU. *Statistical Dynamics: Matter Out of Equilibrium*. Imperial College Press, London, 1997.

[Bal08]   V. BALADI. Linear response despite critical points. *Nonlinearity* **21**, no. 6, (2008), T81–T90.

[BL07]   O. BUTTERLEY and C. LIVERANI. Smooth Anosov flows: correlation spectra and stability. *J. Mod. Dyn.* **1**, no. 2, (2007), 301–322.

[BS08]   V. BALADI and D. SMANIA. Linear response formula for piecewise expanding unimodal maps. *Nonlinearity* **21**, no. 4, (2008), 677–711.

[BY93]   V. BALADI and L.-S. YOUNG. On the spectra of randomly perturbed expanding maps. *Comm. Math. Phys.* **156**, no. 2, (1993), 355–385.

[CE83]   A. P. CARVERHILL and K. D. ELWORTHY. Flows of stochastic dynamical systems: the functional analytic approach. *Z. Wahrsch. Verw. Gebiete* **65**, no. 2, (1983), 245–267.

[Dol04]   D. DOLGOPYAT. On differentiability of SRB states for partially hyperbolic systems. *Invent. Math.* **155**, no. 2, (2004), 389–449.

[DPZ96]   G. DA PRATO and J. ZABCZYK. *Ergodicity for infinite-dimensional systems*, vol. 229 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1996.

[EM01]   W. E and J. C. MATTINGLY. Ergodicity for the Navier-Stokes equation with degenerate random forcing: finite-dimensional approximation. *Comm. Pure Appl. Math.* **54**, no. 11, (2001), 1386–1402.

[GB07]   A. GRITSUN and G. BRANSTATOR. Climate response using a three-dimensional operator based on the fluctuation-dissipation theorem. *J. Atmos. Sci.* **64**, no. 7, (2007), 2558–2575.

[GBD02]   A. S. GRITSOUN, G. BRANSTATOR, and V. P. DYMNIKOV. Construction of the linear response operator of an atmospheric general circulation model to small external forcing. *Russian J. Numer. Anal. Math. Modelling* **17**, no. 5, (2002), 399–416.

[GBM08]   A. GRITSUN, G. BRANSTATOR, and A. MAJDA. Climate response of linear and quadratic functionals using the fluctuation-dissipation theorem. *J. Atmos. Sci.* **65**, no. 9, (2008), 2824–2841.

[GD99]   A. GRITSUN and V. DYMNIKOV. Barotropic atmosphere response to small external actions: Theory and numerical experiments. *Izv. Akad. Nauk. Fiz. Atmos. Okean. Biol.* **35**, no. 5, (1999), 565–581.

[GL06]   S. GOUËZEL and C. LIVERANI. Banach spaces adapted to Anosov systems. *Ergodic Theory Dynam. Systems* **26**, no. 1, (2006), 189–217.

[GL08]   S. GOUËZEL and C. LIVERANI. Compact locally maximal hyperbolic sets for smooth maps: fine statistical properties. *J. Differential Geom.* **79**, no. 3, (2008), 433–477.

[HM06]   M. HAIRER and J. C. MATTINGLY. Ergodicity of the 2D Navier-Stokes equations with degenerate stochastic forcing. *Ann. of Math. (2)* **164**, no. 3, (2006), 993–1032.

[HM08]   M. HAIRER and J. C. MATTINGLY. Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations. *Ann. Probab.* **36**, no. 6, (2008), 2050–2091.

[HM09a]   M. HAIRER and J. C. MATTINGLY. A theory of hypoellipticity and unique ergodicity for semilinear stochastic PDEs, 2009. Preprint.

[HM09b] M. HAIRER and J. C. MATTINGLY. Yet another look at Harris' ergodic theorem for Markov chains, 2009. Preprint.

[HMS09] M. HAIRER, J. C. MATTINGLY, and M. SCHEUTZOW. Asymptotic coupling and a weak form of Harris' theorem with applications to stochastic delay equations, 2009. Prob. Theo. Rel. Fields, to appear.

[JK85] V. JURDJEVIC and I. KUPKA. Polynomial control systems. *Math. Ann.* **272**, no. 3, (1985), 361–368.

[Kat76] T. KATO. *Perturbation theory for linear operators*. Springer-Verlag, Berlin, second ed., 1976. Grundlehren der Mathematischen Wissenschaften, Band 132.

[Kaw90] M. KAWSKI. High-order small-time local controllability. In *Nonlinear controllability and optimal control*, vol. 133 of *Monogr. Textbooks Pure Appl. Math.*, 431–467. Dekker, New York, 1990.

[KL99] G. KELLER and C. LIVERANI. Stability of the spectrum for transfer operators. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)* **28**, no. 1, (1999), 141–152.

[KTH85] R. KUBO, M. TODA, and N. HASHITSUME. *Statistical physics. II*, vol. 31 of *Springer Series in Solid-State Sciences*. Springer-Verlag, Berlin, 1985. Nonequilibrium statistical mechanics, Translated from the Japanese by the authors and Nobuhiko Saitô.

[Kun90] H. KUNITA. *Stochastic flows and stochastic differential equations*, vol. 24 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1990.

[Lei75] C. LEITH. Climate response and fluctuation dissipation. *J. Atmos. Sci.* **32**, no. 20, (1975), 2022–2026.

[MAG05] A. J. MAJDA, R. V. ABRAMOV, and M. J. GROTE. *Information theory and stochastics for multiscale nonlinear systems*, vol. 25 of *CRM Monograph Series*. American Mathematical Society, Providence, RI, 2005.

[MAG10] A. J. MAJDA, R. ABRAMOV, and B. GERSHGORIN. High skill in low-frequency climate response through fluctuation dissipation theorems despite structural instability. *Proc. Nat. Ac. Sci.* **107**, no. 2, (2010), 581–586.

[MBB03] J. MELODY, T. BAŞAR, and F. BULLO. On nonlinear controllability of homogeneous systems linear in control. *IEEE Trans. Automat. Control* **48**, no. 1, (2003), 139–143.

[MFK08] A. J. MAJDA, C. FRANZKE, and B. KHOUIDER. An applied mathematics perspective on stochastic modelling for climate. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **366**, no. 1875, (2008), 2429–2455.

[MGY10] A. MAJDA, B. GERSHGORIN, and Y. YUAN. Low frequency climate response and fluctuation-dissipation theorems: Theory and practice, 2010. Comm. Math. Sci., in press.

[MT93] S. P. MEYN and R. L. TWEEDIE. *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993.

[MW10] A. MAJDA and X. WANG. Linear response theory for statistical ensembles in complex systems with time periodic forcing. *Comm. Math. Sci.* **8**, no. 1, (2010), 142–172.

[Nor86] J. NORRIS. Simplified Malliavin calculus. In *Séminaire de Probabilités, XX, 1984/85*, vol. 1204 of *Lecture Notes in Math.*, 101–130. Springer, Berlin, 1986.

[Nua06] D. NUALART. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, Berlin, second ed., 2006.

[Pal01] T. PALMER. A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Q. J. R. Meteorol. Soc.* **127**, no. 572, Part B, (2001), 279–304.

[PW10]   T. PALMER and P. WILLIAMS, eds. *Stochastic Physics and Climate Modelling*. Cambridge University Press, 2010.

[Ris89]   H. RISKEN. *The Fokker-Planck equation*, vol. 18 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, second ed., 1989. Methods of solution and applications.

[Rom04]   M. ROMITO. Ergodicity of the finite dimensional approximation of the 3D Navier-Stokes equations forced by a degenerate noise. *J. Statist. Phys.* **114**, no. 1-2, (2004), 155–177.

[Rue89]   D. RUELLE. The thermodynamic formalism for expanding maps. *Comm. Math. Phys.* **125**, no. 2, (1989), 239–262.

[Rue97]   D. RUELLE. Differentiation of SRB states. *Comm. Math. Phys.* **187**, no. 1, (1997), 227–241.

[Sus87]   H. J. SUSSMANN. A general theorem on local controllability. *SIAM J. Control Optim.* **25**, no. 1, (1987), 158–194.

[SV79]   D. W. STROOCK and S. R. S. VARADHAN. *Multidimensional diffusion processes*, vol. 233 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1979.

[Vil03]   C. VILLANI. *Topics in optimal transportation*, vol. 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.